

Towards a genetic-based classification of human lung cancer

Iver Petersen* and Simone Petersen

*Institute of Pathology, University Hospital Charité,
Berlin, Germany*

Received 4 October 2000

Accepted 11 December 2000

Lung cancer is a highly aggressive neoplasm which is reflected by a multitude of genetic aberrations being detectable on the chromosomal and molecular level. In order to understand this seemingly genetic chaos, we performed Comparative Genomic Hybridisation (CGH) in a large collective of human lung carcinomas investigating different tumor entities as well as multiple individual tumour specimens of single patients. Despite the considerable genetic instability being reflected by the well known morphological heterogeneity of lung cancer the comparison of different tumour groups using custom made computer software revealed recurrent aberration patterns and highlighted chromosomal imbalances that were significantly associated with morphological histotypes and biological phenotypes. Specifically we identified imbalances in NSCLC being associated with metastasis formation which are typically present in SCLC thus explaining why the latter is such an aggressive neoplasm characterized by widespread tumor dissemination. Based on the genetic data a new model for the development of SCLC is presented. It suggests that SCLC evolving from the same stem cell as NSCLC should be differentiated into primary and secondary tumors. Primary SCLC corresponding to the classical type evolved directly from an epithelial precursor cell. In contrast, secondary SCLC correlating with the combined SCLC develops via an NSCLC intermediate. In addition, we established libraries of differentially expressed genes from different human lung cancer types to identify new candidate genes for several of the chromosomal subregions identified by CGH. In this review, we summarise the status of our results aiming at a refined classification of lung cancer based on the pattern of genetic aberrations.

*Corresponding author: Iver Petersen, MD, Institute of Pathology, Charité, Humboldt-University, Schumannstrasse 20-21, D-10098 Berlin, Germany. Tel.: +49 30 2802 2611; Fax: +49 30 2802 3407; E-mail: iver.petersen@charite.de.

1. Introduction: Concepts in current classification

Lung cancer represents the most important chemically induced tumour type in man since the vast majority of lung cancer are associated with long term cigarette smoking which is true for all major subtypes. The latency is several decades and most carcinomas develop after the age of 50 with the highest incidence around the age of 60. Although there is probably a subgroup of smokers who better tolerate the carcinogen the risk is steadily rising with age and still many patients become symptomatic within the 70-ties or 80-ties [9].

The major classification schemes of lung cancer are represented in Table 1. The new WHO classification defines 9 subgroups of malignant epithelial lung tumours of which the major subcategories are squamous cell carcinoma, adenocarcinoma, large cell carcinoma and small cell carcinoma [32]. The authors also present the spectrum of neuroendocrine lung tumours ranging from the typical carcinoid, the atypical carcinoid to the highly aggressive neuroendocrine carcinomas consisting mainly of LCNEC and SCLC. This concept suggests that there is a transition from the more benign carcinoid tumours to the highly malignant SCLC. In other classifications this has lead to the designation of neuroendocrine tumors (NET) for the benign tumours and neuroendocrine carcinoma (NEC) for the malignant ones. It still seems to be based on the old assumption that SCLC and LCNEC are derived from a neuroendocrine precursor cell, i.e., the Kulchitsky cell, whereas NSCLC develop from an epithelial precursor cell. However, these transitions are very rarely observed. Instead there is evidence that SCLC and NSCLC may be derived from the same cell type, sometimes called the amphicrine stem cell since it may give rise to a neuroendocrine as well as epithelial differentiation [10].

Whereas the major subtypes adenocarcinoma, SCC, LCC and SCLC are relatively easy recognisable it is fairly difficult to differentiate some variants from each other. For instance, many pathologist will have diffi-

Table 1
Classification schemes of malignant epithelial lung tumors

Actual WHO classification
1. Squamous cell carcinoma (SCC) Variants: papillary, clear cell, small cell, basaloid SCC
2. Small cell carcinoma (SCLC) Variant: combined SCLC
3. Adenocarcinoma (Adeno) Variants: acinar, papillary, bronchioloalveolar, solid, mixed, others
4. Large cell carcinoma (LCC) Variants: LCNEC, basaloid, lymphoepithelioma-like, clear cell, with rhabdoid phenotype
5. Adenosquamous carcinoma
6. Carcinoma with pleomorphic, sarcomatoid or sarcomatous elements
7. Carcinoid tumors Variants: typical, atypical
8. Carcinomas of salivary-gland type
9. Unclassified carcinomas
Spectrum of neuroendocrine lung tumors Typical carcinoid – Atypical carcinoid – LCNEC – SCLC
Major clinical differentiation NSCLC (Adeno, SCC, LCC) ↔ SCLC

Abbreviations: LCNEC, large cell neuroendocrine carcinoma, NSCLC, non-small cell lung carcinoma (other abbreviations as mentioned in the table).

culties to subdivide large cell carcinomas into the four possible subgroups, i.e., classical LCC lacking neuroendocrine morphology and neuroendocrine differentiation accessed by immunohistochemistry and/or electron microscopy, LCC with neuroendocrine differentiation lacking neuroendocrine morphology, LCC only with neuroendocrine morphology and finally LCNEC showing both characteristics.

Although not explicitly stated in the WHO classification SCLC actually consists of two variants, one might be termed the classical type showing either the typical oat cell or fusiform morphology and the combined SCLC presenting as a mixture of an NSCLC and SCLC. In the clinical setting the major distinction is between SCLC (~20% of all lung cancers) and NSCLC (~80%). The latter may show either the pure differentiation of one of the major types (SCC, Adeno, LCC) or a mixed differentiation which is a frequent finding due to the high percentage of morphologically heterogeneous tumours [11]. The histopathological diagnosis of SCLC is highly predictable for the clinical course which is characterised by a highly aggressive phenotype with early and widespread tumor dissemination and excellent response to chemotherapy in the

initial phase. In contrast, it is far more difficult to predict the outcome of NSCLC based on morphological grounds. There are tumours similar to SCLC with a highly aggressive behaviour and early metastasis formation, whereas others may be cured after surgery or remain stable for a considerable period of time even in the case of residual disease.

In this review we describe our strategy and results in the attempt to contribute to the understanding of human lung cancer genetics and why we believe that this knowledge will have a major impact on a refined classification of the disease.

2. CGH results

Comparative Genomic Hybridization (CGH) is a molecular cytogenetic technique for the detection of DNA gains and losses [5]. As a screening method it provides a survey of the entire tumor genome. DNA overrepresentations are potentially associated with the activation of proto-oncogenes while deletions might indicate the inactivation of a tumor suppressor gene. For conventional CGH, however, the resolution is limited to approximately one chromosomal band allowing the detection of deletions in the order of 10 Mb whereas for amplifications a size of 2 Mb (product of amplicon size and copy number) is achievable [1]. Thus any correlation with the status of a gene within a specific chromosomal region needs to be confirmed by additional studies. Recently new approaches have been published which will extend the resolution of CGH [24,30]. However, the majority of CGH data now arising in the literature is still gathered by the classical techniques using normal chromosome spreads as the DNA matrix to which the genomes bind.

In the recent years we analyzed a collective of lung carcinomas by CGH comprising more than 250 tumour specimens. The data has either been published [13, 14,18,29,33] or is available at our CGH online tumor database at <http://amba.charite.de/cgh>. Beside primary tumors also metastases and tumor cell lines were analyzed. The tumour DNA were mainly obtained from frozen tissue derived from surgical resections at the Department of Surgery of the Charité Hospital at the Humboldt-University Berlin. Additionally, snap frozen tumour specimens of primary and metastatic lesions were collected at post mortem examinations.

Based on our own CGH software [27] we extended the functionality of conventional CGH programs by

taking advantage of the fact that the primary data in CGH being derived from digital analysis of fluorescence images is already computerised to calculate histograms and difference histograms [14]. There are two major advantages of this approach. First the number of tumour samples that can be simultaneously visualised and analysed is virtually unlimited. Second and even more important is the fact that allows the statistical comparison of tumour subgroups highlighting distinct chromosomal subregions of which the difference in gains or losses between these subgroups is statistically significant.

The importance of the DNA gains and losses detected by CGH are underlined by the fact that karyotyping and more recently M-FISH and SKY analysis failed to detect recurrent translocations in lung cancer [31]. Thus in contrast to leukaemia, lymphoma and sarcoma, oncogenic fusion protein do not seem to play an important role in solid tumours.

2.1. Comparison of SCLC and NSCLC

Figure 1A shows the histogram of SCLC. The SCLC collective comprised the previously published autopsy cases [13,26,29] and several additional samples including 7 tumour cell lines. In general, the results of the autopsy cases were similar to those of the cell lines suggesting that SCLC cell lines constitute a valuable model to study this tumour type. The typical findings in SCLC are deletions on chromosomes 3p, 4, 5q, 10q, 13q, 17p and DNA gains on 3q, 5p, 6p, 8q and 17q. Interestingly, deletions are more frequent than DNA gains suggesting that the inactivation of tumour suppressor genes is as important as gain of function mutations of proto-oncogenes.

SCLC usually harbour large deletions affecting entire chromosome arms or whole chromosomes as exemplified by the deletions on 3p, 10 and 17p. Deletions of 3p have been reported as the characteristic finding in SCLC [35]. Since other tumour types also carry a high incidence of DNA losses on 3p the deletion per se can not be considered specific. However, there is a typical pattern regarding chromosome 3 in SCLC which consists of the deletion of the entire short chromosome along with the overrepresentation of the long arm, i.e., the CGH equivalent of a 3q isochromosome. The 3p deletion and the 3q gain often fulfil the criteria of pronounced imbalances being defined as those imbalances with a ratio exceeding the thresholds of 0.5 and 1.5 [18]. The pronounced losses (ratio < 0.5) and gains (ratio > 1.5) are related to multi copy dele-

tions or high copy amplifications, respectively. These alterations are frequently observed in SCLC and correspond to the observation that the tumour often harbours amplifications [3].

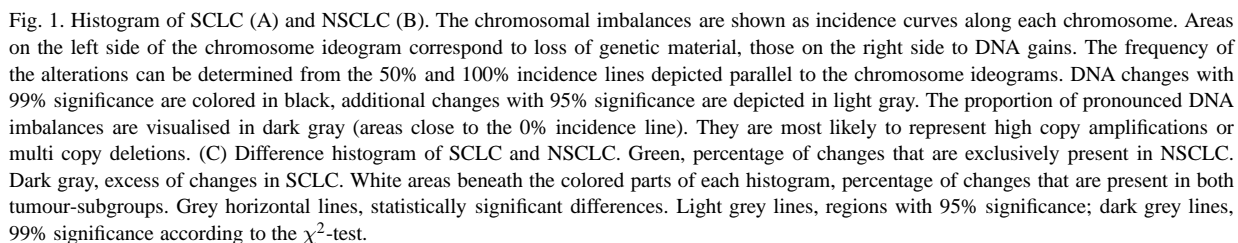
Our relatively small study on neuroendocrine lung tumors showed that LCNEC has similar alterations as SCLC, particularly the deletions of 3p and 10q were frequently found. Interestingly, the CGH patterns of carcinoids and neuroendocrine carcinomas were not similar [33].

NSCLC showed overlapping as well as different alterations compared to SCLC (Fig. 1B). They also carried a high incidence of deletions on chromosomes 3p, 4, 5 and 13q. In addition DNA gains occurred at a high frequency on chromosomes 1p, 6q, 9p, 18q and 21q. Also for the common chromosomes the patterns are slightly different. For instance, chromosome 3p is particularly affected by interstitial deletions whereas in SCLC the entire chromosome arm or large regions of it are affected. For chromosome 13q, SCLC showed deletions of the proximal arm including the locus of the Rb gene at 13q14 whereas in NSCLC typically the distal chromosome arm is lost.

The above mentioned differences are best visualised by the difference histogram between SCLC and NSCLC shown in Fig. 1C. It clearly indicates that the deletions of the entire chromosome 3p, 10, 4p16, 15q, 16q, 17p as well as the overrepresentations on chromosomes 1, 3q, 6, 13, and 17q24–q25 are significantly associated with SCLC. In contrast, the deletions of chromosome 1p, 6q, 9p, 18q, 21q and the gain of chromosome 22q is significantly more frequent in NSCLC.

2.2. Comparison of non-metastatic and metastatic NSCLC

In a recent study we examined primary SCC without evidence of tumor dissemination, i.e., stage pN0 and pM0, with metastasising carcinomas showing hematogenous (pM1) and/or lymphatic tumour spread (pN+). For most chromosomal regions the latter tumour group harboured more alterations which is consistent the paradigm of tumour genetics postulating that tumour progression and metastasis formation is characterised by an accumulation of genetic defects. Specifically, the deletions at 3p12–p14, 4p15–p16, and 10q as well as the gain on chromosome 1q22–q25 were associated with the metastatic phenotype [18]. In another study of 42 brain metastases we additionally observed a peak in the histogram for the gain at 17q24–



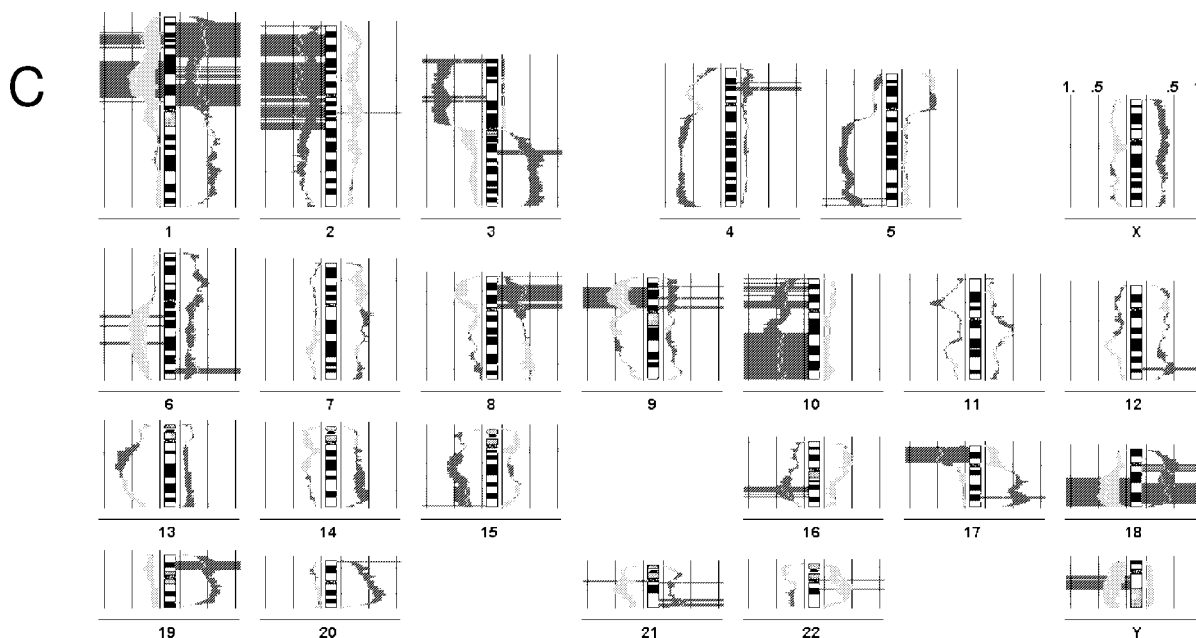


Fig. 1. (Continued).

q25 suggesting that the amplification of a gene at this chromosomal regions might mediate tumour dissemination into the nervous system [19]. Interestingly, these alterations were also associated with the SCLC phenotype providing a genetic correlate to the fact that SCLC is a highly metastatic tumour type which often spreads into the brain.

The statistical analysis of the tumour subgroups (pM0/pN0 SCC versus pM1/pN+ SCC) was supplemented by the analysis of primary and corresponding metastatic tumours. Although the tumours of the same patients harboured a high percentage of common imbalances indicating the clonal relationship the comparison also revealed a considerable heterogeneity indicating the genetic instability of lung cancer on the chromosomal level. Individual imbalances that were indicated by the statistical analysis of the tumour groups could be additionally found in metastatic lesions. However, many of them were already detectable the primary tumour similarly to our findings in primary and metastatic SCLC [29]. There are two main conclusions from this. First, the comparison of tumour subgroups seems more appropriate to dissect the genetic alterations that are responsible for a tumour phenotype than the comparison of individual tumours since the latter type of analysis is largely biased by the genetic tumour heterogeneity. Second and even more importantly, the potential of a tumour for hematogeneous spread can at least partially be deduced by the genetic analysis of a

primary tumour. This offers an important outlook for a refined new classification since small biopsies might be used to determine the malignant potential. Similar to morphological grading such a genetic grading must be correlated in its predictive value to conventional tumour staging which still constitutes the gold standard to describe the status quo of the biological tumour phenotype. However, our recent analysis of primary head and neck squamous cell carcinomas, the first tumor collective where patients survival data was available, indicated that the analysis of the chromosomal imbalances are even a better prognostic indicator than pTNM staging [2]. This is a very encouraging result.

Finally the comparison of the primary SCC with their corresponding metastases also suggested certain mechanisms of the clonal evolution of chromosomal changes during tumour progression. Overrepresentations were reduced in size. In particular, this was observed for gain of chromosome 1q which resulted in the overrepresentation of the centromeric region 1q21–q25 constituting exactly the region putatively harbouring a proto-oncogene of relevance in metastasis formation. In contrast, small interstitial deletions were often extended to the loss of entire chromosome arms or even whole chromosomes which was seen for several regions, e.g., chromosome 10q [18].

The analysis of different types of NSCLC indicated chromosomal imbalances that were associated with tumour differentiation. Adenocarcinomas, for instance,

are typically characterised by overrepresentations of chromosome 1q. This again provides a genetic correlate to the fact that adenocarcinoma carry a higher risk for hematogenous metastasis formation than lung SCC [14].

2.3. Morphological tumour heterogeneity: analysis of a combined SCLC

A metastasizing combined SCLC was analysed after microdissection. The primary tumour showed a squamous cell differentiation together with a SCLC component (Fig. 2A). A synchronous lung metastasis showed exclusively the SCLC phenotype (Fig. 2B). Not surprisingly, CGH revealed a clonal relationship between both tumours as shown in Fig. 2C and 2D, respectively. Thus, the SCLC must have evolved from the SCC component of the tumour. It is interesting to note, that the expression of the neuroendocrine marker Synaptophysin and Chromogranin was restricted to the SCLC component while the SCC was negative.

3. Model of primary and secondary SCLC

The above mentioned CGH results indicate that SCLC and NSCLC are genetically related tumour entities and that NSCLC have the potential to evolve into a SCLC during tumour progression. Obviously the two components of combined SCLC are not genetically independent. Rather the morphological heterogeneity is a detectable correlation to the genetic instability of this tumour type and lung cancer in general. A model for the development of SCLC which is based on the overlapping CGH patterns of metastatic SCC and SCLC, the chromosomal mechanisms during tumour progression as well as the analysis of the combined SCLC is depicted in Fig. 3. The primary SCLC correlating with the classical SCLC develops directly from a precursor cell of probably epithelial origin. In contrast, the secondary SCLC evolve via a NSCLC intermediate. The secondary SCLC thus correlates with the combined SCLC. It is important to note that it is also a model of lung cancer in general in which small cell and neuroendocrine differentiation should be considered as markers for tumor progression. In contrast we

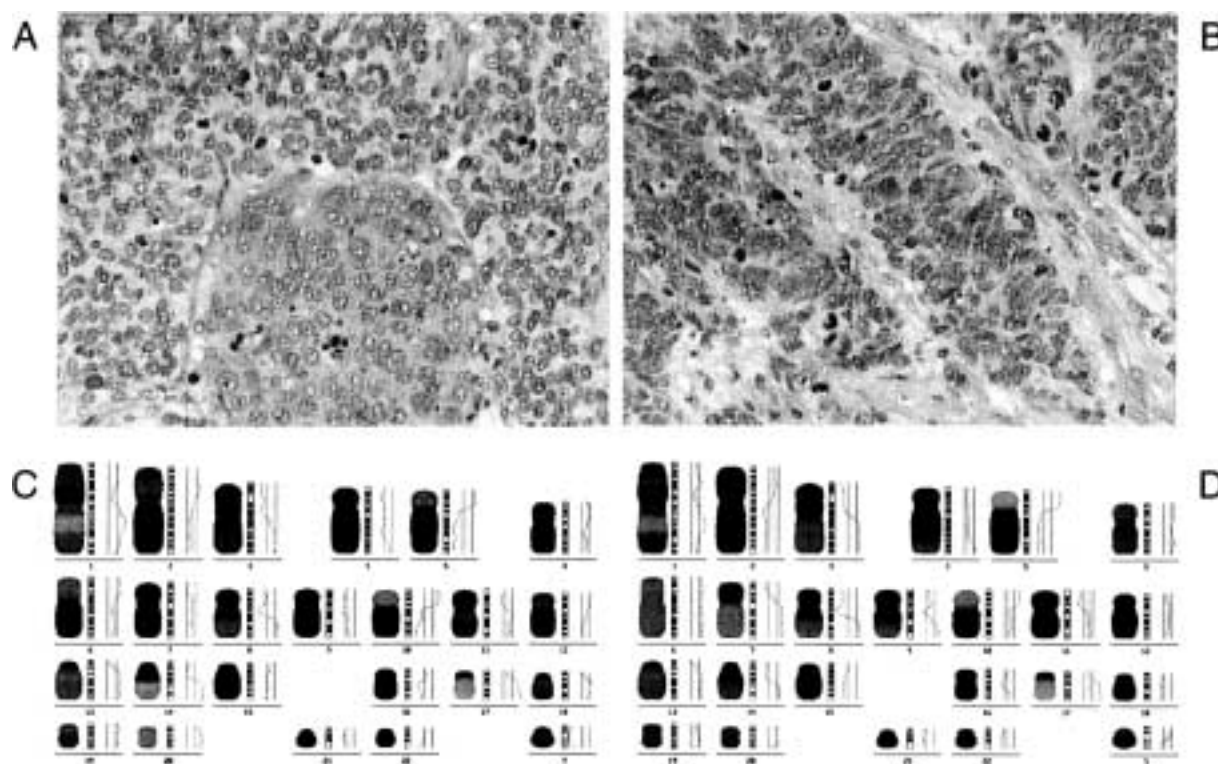


Fig. 2. H&E stains of the primary combined SCLC showing of a SCC component (A) and a synchronous metastasis exclusively with the SCLC phenotype (B). After microdissection the SCC component of the primary tumour (C) and the SCLC metastasis (D) was analysed by CGH showing a clonal relationship as evidenced by the high number of common changes.

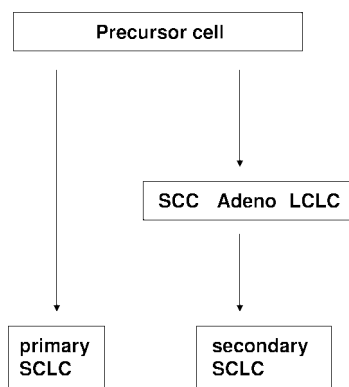


Fig. 3. Model of primary and secondary SCLC.

strongly feel that the presence of neuroendocrine proteins should not be used as an indicator of a putative tumour stem. Of course, these consideration apply only for lung carcinomas and not for carcinoids or benign neuroendocrine tumours of the lung.

The model strengthen the notion that NSCLC and SCLC are derived from the same stem cell and that both tumour types are more closely related to each other than SCLC with other neuroendocrine lung tumors, in particular carcinoids. Epidemiology and pathology also support this view. Cigarette smoking is the typical risk factor for SCLC and similar to NSCLC, dysplasia and preneoplastic epithelial changes are a common finding. In contrast, neuroendocrine precursor lesions are only very rarely observed. Thus, for the majority of cases the neuroendocrine carcinomas SCLC and LCNEC should be regarded as distinct tumor types than carcinoids being derived from different stem cell. Although there may exist the progression of a typical to an atypical carcinoid and finally a SCLC, the progression of a NSCLC to a SCLC is probably much more frequent. Therefore, SCLC might be regarded as an end stage tumour with the most aggressive phenotype of all lung carcinomas. The situation is similar to glioblastoma multiforme of the brain. Mostly it primarily presents as a glioblastoma but in some cases first an astrocytoma is diagnosed which later progress to the most malignant brain tumor in man. Interestingly, glioblastomas and SCLC also have some chromosomal alterations in common, in particular the loss of chromosome 10 [34].

4. Correlation between chromosomal changes and genetic defects

For chromosome 10q, we performed extensive additional genetic analysis which showed an excellent

correlation with the CGH data. Using allelotyping, we confirmed that deletions of this chromosome arm are a typical finding in SCLC and that it is associated with tumour progression and metastasis formation of lung SCC. We identified three minimal regions of deletions putatively harbouring the tumour suppressor genes. However, any of three major candidate genes, i.e., *MXI1*, *PTEN/MMAC1* and *DMBT1*, were not affected in lung cancer [15,16,20].

Our CGH results is also in excellent agreement with the data on the prevalence of specific gene defects in different lung cancer types published by others and us. For instance, SCLC typically show a high incidence of deletions at 17p13 and 13q14 correlating with the observations that *TP53* and *RB1* are frequently inactivated in this tumour type. In NSCLC, the high incidence of DNA gains at 11q13 reflects the fact that cyclin D1 gene is frequently amplified [6,25,28,36].

In summary, the CGH data clearly indicate the feasibility of a genetic lung tumour classification with the potential to provide superior results to morphological characterisation, in particular for the possible assessment of the metastatic potential.

There are however some disadvantages that will probably prevent CGH from becoming a routine method in genetic tumour classification. As mentioned above, the method has a limited resolution. In addition, there is a considerable work load associated with the analysis of a single tumour [12]. First the preparative steps, i.e., DNA extraction, labeling and hybridization, take several days. Second, image capture and analysis and in particular the careful karyotyping of the 15 metaphases that we usually analyse takes several hours. Thus, in an optimistic estimate the final result will be available after one week with one sample per day and technician. In comparison, conventional processing of the tumour specimen and histopathological evaluation is much faster and cheaper. Therefore it is an important question whether a laborious CGH analysis can be replaced by other methods.

Copy number changes and loss or gain of function mutations are frequently associated with either reduced expression or overexpression of tumor associated genes. We therefore performed an immunohistochemical analysis of the *HER/NEU* proto-oncogene to answer the question whether the DNA gain on chromosome 17q21 where the gene is located correlates with protein overexpression [8]. As depicted in Fig. 4, this was indeed the case. Similar to breast carcinomas the laborious genetic analysis can thus be largely replaced by a simple immunohistochemical test.

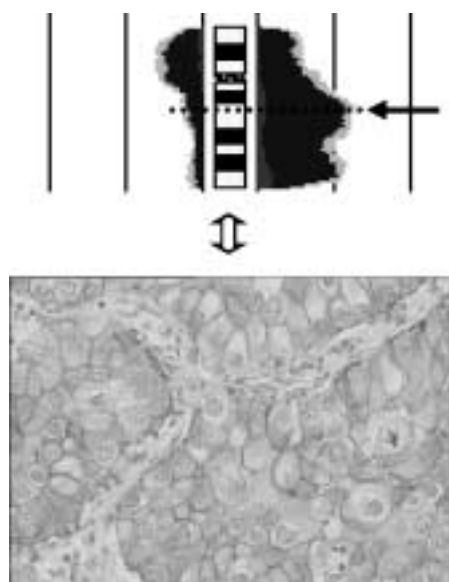


Fig. 4. The overexpression of HER2/NEU being located at chromosome 17q21 correlates with DNA gains of the respective chromosomal band in NSCLC.

Thus, the chromosomal imbalances are most probably associated with a distinct pattern of gene expression mediating the biological phenotype of the tumor. This hypothesis is strongly supported by recent cDNA microarray analysis of breast carcinomas [23].

Based on our negative experience with the candidate genes on chromosome 10q and taking into consideration that many tumour associated genes are either not yet identified or fully characterised we sought for alternative methods to rapidly identify new candidate genes in lung cancer.

5. Identification and characterisation of new candidate genes

5.1. Generation of cDNA libraries of differentially expressed genes

Abnormal gene expression is a hallmark of the neoplastic phenotype. Therefore we used expression genetics to identify new candidate genes in lung carcinogenesis. Specifically we applied the method Subtractive Suppression Hybridization (SSH) to generate libraries of differentially expressed genes [4]. Without going into detail the technique is based on the subtraction hybridisation of so called tester cDNA versus a driver cDNA which results in the enrichment of clones of either cDNA pools. By using the tumour first as the

Table 2
Results of the SSH analysis

Libraries (SCLC, SCC, Adeno)	6
Total number of clones	2471
Already sequenced	1554
Known genes	49% (40–60%)
Expressed sequence tags	39% (24–60%)
Unknown sequences	12% (10–16%)
Analyzed by Northern blot	363
Differentially expressed	73.7% (54–92%)

driver and then as the tester two libraries are generated that will represent mainly the genes that are overexpressed and underexpressed within the tumour, respectively. We meanwhile generated six of such libraries by comparing an adenocarcinoma, a SCLC and a SCC with normal bronchial epithelial cells [21]. A compilation of the data is given in Table 2.

It is important to note that more than 10% of the clones are not represented in the public databases. About 40% correspond to so called expressed sequence tags which are cDNA fragments of genes of which the full length cDNA has not yet been determined and the function is unknown. Although approximately half of the clones represent known genes being listed in the Unigene set of the NCBI database their importance in tumorigenesis of lung cancer is mostly unknown.

We are performing Northern blot analysis to confirm the expression pattern of the cDNA clones [21]. Although this technique is laborious and might appear somehow old-fashioned in view of the modern screening methods like cDNA arrays it has several important advantages. First, Northern blotting is still the gold standard for the assessment of gene expression on the RNA level. For our libraries, it indicated that more than 70% of the clones are indeed differentially expressed. Second, it provides a measurement for the size of the full length RNA transcripts which is essential to know if the entire cDNA needs to be isolated. Third, by including additional RNA samples in the Northern blot analysis the frequency of either reduced or enhanced gene overexpression in the tumor type can be estimated. We meanwhile analyze at least RNA samples from 3 tumor cell lines from a SCLC, SCC and adenocarcinoma by this procedure. Although it is too early to draw a final conclusion on the patterns of gene expression in SCLC and NSCLC the fact that several genes are represented in the different libraries and that many show a similar expression pattern in the Northern blot analysis supports the above mentioned model of primary and secondary SCLC.

Table 3
Overexpressed genes identified by SSH

Protein type	Protein name
Oncogenes	N-myc, c-myc, Ki-ras2, N-ras-related gene unr
Receptor tyrosine kinases	Epidermal growth factor receptor EGFR, homolog to the ERBB2 receptor tyrosine kinase
SH ₂ -domain proteins	Grb7, Grb14
Cell-cell adhesion molecules	Intercellular adhesion molecule I ICAM-I
Cytosolic proteins	Stathmin Op18, Ran-binding proteins 1 and 5, centrosome-associated kinase BTAK
Growth factors/cytokines	Glia maturation factor β GMFB, interleukin 8
Cytoskeletal proteins	Zytokeratin K18, actin-bundling protein L-plastin
DNA interaction	Topoisomerase 1 and 2 α
Regulatory proteins	E1A binding protein p300, translation initiation factor INT6, replication factor C
Surface molecules, antigens	Epithelial membrane protein 1 EMP1, carcinoembryonic antigen 1 CEA1, NY-CO-1, NY-CO-25, epithelial glycoprotein EGP
Ca ⁺⁺ -dependent enzymes	Calpain, calyculin-binding protein
Energy metabolism	Cytochrome <i>c</i> , cytochrome <i>b</i> , mitochondrial proteins

Table 4
Underexpressed genes identified by SSH

Protein type	Protein name
Cell adhesion and communication	
Extracellular matrix proteins	Laminin α 3a, fibronectin, thrombospondin 1
Cell membrane associated proteins	CD44-variant, integrin α 6, integrin β 1, connexin 26, tissue factor, desmocollin Typ4, fibronectinreceptor, bullous pemphigoid antigen 1, desmoplakin1 and 2, oncostatin M-receptor
Protease inhibitors	Cystatin A, maspin, plasminogen-activator-inhibitor 2, VATKI, SSCA1 epidermal differentiation complex, small proline-rich proteins 1 and 3, cornifin B
Ca ⁺⁺ -binding proteins	CaN19, MRP-8/calgranulin A, calgranulin B, Ca ⁺⁺ -activated chloride channel protein 2 and 3 (CaCC2,3)
Cytoskeleton protein	
Cytokeratins	K6A, K6, K15
Others	β - γ -non-lens-crystallin AIM1, moesin ERM-binding phosphoprotein 50
Regulatory proteins	
Transcription factors	p51B/p73H, CUSP (member of the p53-family)

The known genes belong to very different classes of proteins, e.g., being involved in cellular signaling, the cytoskeleton and the interaction with the extracellular matrix. Some examples of these genes are listed in Tables 3 and 4.

5.2. Characterization of the human Calyculin binding protein

We meanwhile fully characterized one gene on chromosome 1q24-q25 [22]. This was greatly facilitated by the fact that a sequenced genomic clone was available. Since the Human Genome Project is finishing its completion, the entire sequence of the human genome will

be soon accessible. Thus our procedure is exemplary for future gene discovery.

Starting from the cDNA fragment of our first library we performed a multi-tissue Northern blotting. According to the expression pattern we ask the Ressource Center of the German Human Genome Project to provide a full length cDNA clone of the gene by screening one such library of an organ in which the gene is strongly expressed. In the meantime we ordered the genomic PAC clone from the Sanger Centre in the UK which performed the genomic sequencing for FISH experiments. After confirming and sequencing of the full length cDNA clone it was fairly easy to reveal the genomic structure of the gene by comparing the cDNA and the genomic sequence. The gene encodes for the

human homologue of the Calcyclin-binding protein. It was mapped to the 1q24–q25 and spans about 10 kb of genomic DNA with a 1.5 kb mature transcript. The putative protein encoded for 228 amino acids and harbors a nuclear localization signal and several protein kinase phosphorylation sites. The gene showed overexpression in the majority of lung cancer cell lines as well as advanced primary tumors. In addition, FISH analysis with a PAC clone at 1q24–q25 covering the genomic sequence indicated intrachromosomal and interchromosomal rearrangements associated with gene amplification. The data suggest that the human calcyclin binding protein acts as a putative proto-oncogene during progression of lung cancer and possibly other advanced tumor types. This data was accumulated in a couple of months in contrast to years needed in before all the above mentioned facilities were available. We are meanwhile in the process of generation antibodies against the gene to test it more easily on clinical tumor samples.

6. Perspective

The most important task in the forthcoming years will be first the identification of all genes involved in lung tumorigenesis. Out of this pool of probably several thousand genes those need to be selected that carry the highest clinical significance. This will require the use of modern screening technologies like cDNA arrays and also sophisticated bioinformatics to analyse the wealth of data. These genes need to be assessed on clinical specimens for which the use of tissue arrays provide a very promising tool [7]. Finally the structure and function of these genes need to be clarified for potential therapeutic intervention. Regarding tumour diagnostics the methodology how these new markers will be investigated in a routine setting will largely depend on the number of genes that need to be surveyed for a genetic classification. If a large number of genes, e.g., more than 50, have to be assessed simultaneously, biomolecules extracted from the tumour tissue probably will be analysed by chip technologies. This, however, will require a change in classical procedure of tissue processing in pathology which needs to be complemented by cryopreservation since most of these technologies require fresh frozen material. However, it is well perceivable that the number of markers remain so small that they might still be analysed by conventional immunohistochemistry one by one. Anyhow, these are very exiting years for tumour diagnostics and pathology is the medical discipline which will provide major impacts and will first benefit from this development.

Acknowledgements

We would like to thank all the technicians, postdoctoral fellows and medical students that contributed to this work, i.e., Nicole Deutschmann, Manuela Pacyna-Gengelbach, Christa Schütze, Jacqueline Rudolf, Cordula Heckert, Günter Wolf, Karsten Schlüns, Karl Roth, Holger Langreck, Glen Kristiansen, Almut Goeze, Sven Schmid, Blend Krebber, Marco Aninat Meyer. We are deeply indebted to our institute director Manfred Dietel for his long term continuous support. We enjoyed the stimulating collaborations with several colleague within and outside the Charité, in particular Ulrike Bockmühl, Klaus Gellert, Thomas Ried and Michael Speicher. The work was supported by the Charité university hospital, Deutsche Forschungsgemeinschaft, Berliner Krebsgesellschaft, Deutsche Krebshilfe and the Monika Kutzner-Stiftung.

References

- [1] M. Bentz, A. Plesch, S. Stilgenbauer, H. Dohner and P. Lichter, Minimal sizes of deletions detected by comparative genomic hybridization, *Genes Chromosomes Cancer* **21** (1998), 172–175.
- [2] U. Bockmühl, K. Schlüns, I. Kuchler, S. Petersen and I. Petersen, Genetic imbalances with impact on survival in head and neck cancer patients, *Am. J. Pathol.* **157** (2000), 369–375.
- [3] O. Brison, Gene amplification and tumor progression, *Biochim. Biophys. Acta* **1155** (1993), 25–41.
- [4] L. Diatchenko, Y.F. Lau, A.P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E.D. Sverdlov and P.D. Siebert, Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries, *Proc. Natl. Acad. Sci. USA* **93** (1996), 6025–6030.
- [5] A. Kallioniemi, O.-P. Kallioniemi, D. Sudar, D. Rutovitz, J.W. Gray, F. Waldman and D. Pinkel, Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors, *Science* **258** (1992), 818–821.
- [6] M.J. Kelley, K. Nakagawa, S.M. Steinberg, J.L. Mulshine, A. Kamb and B.E. Johnson, Differential inactivation of CDKN2 and Rb protein in non-small-cell and small-cell lung cancer cell lines, *J. Natl. Cancer Inst.* **87** (1995), 756–761.
- [7] J. Kononen, L. Bubendorf, A. Kallioniemi, M. Barlund, P. Schraml, S. Leighton, J. Torhorst, M.J. Mihatsch, G. Sauter and O.P. Kallioniemi, Tissue microarrays for high-throughput molecular profiling of tumor specimens, *Nat. Med.* **4** (1998), 844–847.
- [8] G. Kristiansen, S. Petersen, O. Kaufmann, K. Schlüns, M. Dietel and I. Petersen, Overexpression of c-erbB2 protein correlates with disease-stage and chromosomal gain at the c-erbB2 locus in non-small cell lung cancer, *European J. Cancer* (2000), in press.

- [9] M.E. Mattson, E.S. Pollack and J.W. Cullen, What are the odds that smoking will kill you?, *Am. J. Public Health* **77** (1987), 425–431.
- [10] K.M. Müller and S. Gonzalez, Preneoplasias and early carcinoma of the lungs – histogenetic aspects of bronchial carcinoma, *Pneumologie* **45** (1991), 971–976.
- [11] K.M. Müller and A. Fisseler-Eckhoff, What's new in lung tumor heterogeneity?, *Path. Res. Pract.* **184** (1989), 108–115.
- [12] I. Petersen, A. Schwendel, U. Bockmühl and M. Dietel, Die Komparative Genomische Hybridisierung – eine Screening-methode in der genetischen Tumordiagnostik, *Pathologie* **17** (1996), 333–341.
- [13] I. Petersen, H. Langreck, G. Wolf, A. Schwendel, R. Psille, P. Vogt, M.B. Reichel, T. Ried and M. Dietel, Small cell lung cancer is characterized by a high incidence of deletions on chromosomes 3p, 4q, 5q, 10q, 13q and 17p, *Br. J. Cancer* **75** (1997), 79–86.
- [14] I. Petersen, M. Bujard, S. Petersen, G. Wolf, A. Goeze, A. Schwendel, H. Langreck, K. Gellert, M. Reichel, K. Just, S. du Manoir, T. Cremer, M. Dietel and T. Ried, Patterns of chromosomal imbalances in adenocarcinoma and squamous cell carcinoma of the lung, *Cancer Res.* **57** (1997), 2331–2335.
- [15] S. Petersen, G. Wolf, U. Bockmühl, K. Gellert, M. Dietel and I. Petersen, Deletions on chromosome 10q in human lung cancer: Association with tumor progression and metastatic phenotype, *Br. J. Cancer* **75** (1998), 79–86.
- [16] S. Petersen, J. Rudolf, U. Bockmühl, K. Gellert, G. Wolf, M. Dietel and I. Petersen, Distinct regions of allelic imbalance on chromosome 10q22–q26 in squamous cell carcinomas of the lung, *Oncogene* **17** (1998), 449–454.
- [17] S. Petersen, G. Wolf, U. Bockmühl, K. Gellert, M. Dietel and I. Petersen, Allelic loss on chromosome 10q in human lung cancer: association with tumor progression and metastatic phenotype, *Br. J. Cancer* **77** (1998), 270–276.
- [18] S. Petersen, M. Aninat-Meyer, K. Schlüns, K. Gellert, M. Dietel and I. Petersen, Chromosomal alterations in the clonal evolution to the metastatic stage of squamous cell carcinomas of the lung, *Br. J. Cancer* **82** (2000), 65–73.
- [19] I. Petersen, H. Hidalgo, S. Petersen, M. Pacyna-Gengelbach, K. Schlüns, A. Goeze, B. Krebber, J. Szymas and A. von Deimling, Chromosomal imbalances in brain metastases of solid tumors, *Brain Pathol.* **10** (2000), 395–401.
- [20] S. Petersen, J. Rudolf, U. Bockmühl, N. Deutschmann, M. Dietel and I. Petersen, Analysis of the DMBT1 gene in carcinomas of the respiratory tract, *Int. J. Cancer* **88** (2000), 71–76.
- [21] S. Petersen, C. Heckert, J. Rudolf, K. Schlüns, O.I. Tchernitsa, R. Schäfer and I. Petersen, Gene expression profiling of advanced lung cancer, *Int. J. Cancer* **86** (2000), 512–517.
- [22] S. Petersen, N. Deutschmann, C. Sers, M. Dietel and I. Petersen, Amplification and overexpression of the gene encoding the human calyculin binding protein on chromosome 1q24–q25 in advanced lung carcinomas (2000), manuscript submitted.
- [23] C.M. Perou, S.S. Jeffrey, M. van de Rijn, C.A. Rees, M.B. Eisen, D.T. Ross, A. Pergamenschikov, C.F. Williams, S.X. Zhu, J.C. Lee, D. Lashkari, D. Shalon, P.O. Brown and D. Botstein, Distinctive gene expression patterns in human mammary epithelial cells and breast cancers, *Proc. Natl. Acad. Sci. USA* **96** (1999), 9212–9217.
- [24] J.R. Pollack, C.M. Perou, A.A. Alizadeh, M.B. Eisen, A. Pergamenschikov, C.F. Williams, S.S. Jeffrey, D. Botstein and P.O. Brown, Genome-wide analysis of DNA copy-number changes using cDNA microarrays, *Nat. Genet.* **23** (1999), 41–46.
- [25] M.B. Reichel, H. Ohgaki, I. Petersen and P. Kleihues, p53 mutations in primary human lung tumors and their metastases, *Molecular Carcinogenesis* **9** (1994), 105–109.
- [26] T. Ried, I. Petersen, H. Holtgreve-Grez, M.R. Speicher, E. Schröck, S. du Manoir and T. Cremer, Mapping of multiple DNA gains and losses in primary small cell lung carcinomas by comparative genomic hybridization, *Cancer Res.* **54** (1994), 1801–1806.
- [27] K. Roth, G. Wolf, M. Dietel and I. Petersen, Image analysis for comparative genomic hybridization based on a karyotyping program for Windows, *Anal. Quant. Cytol. Histol.* **19** (1997), 461–474.
- [28] Y. Sameshima, Y. Matsuno, S. Hirohashi, Y. Shimosato, H. Mizoguchi, T. Sugimura, M. Terada and J. Yokota, Alterations of the p53 gene are common and critical events for the maintenance of malignant phenotypes in small-cell lung carcinoma, *Oncogene* **7** (1992), 451–457.
- [29] A. Schwendel, H. Langreck, M. Reichel, E. Schröck, T. Ried, M. Dietel and I. Petersen, Primary small cell lung carcinomas and their metastases are characterized by a recurrent pattern of genetic alterations, *Int. J. Cancer (Pred. Oncol.)* **74** (1997), 86–93.
- [30] S. Solinas-Toldo, S. Lampel, S. Stilgenbauer, J. Nickolenko, A. Benner, H. Dohner, T. Cremer and P. Lichter, Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances, *Genes Chromosomes Cancer* **20** (1997), 399–407.
- [31] M.R. Speicher, S. Petersen, S. Uhrig, I. Jentsch and I. Petersen, Analysis of karyotypic instability in non-small cell lung cancer by multiplex-FISH (M-FISH) and comparative genomic hybridization (CGH), *Lab. Invest.* **80** (2000), 1031–1041.
- [32] W.D. Travis, T.V. Colby, B. Corrin, Y. Shimosato and E. Brambilla, *WHO International Histological Classification of Tumours: Histological Typing of Lung and Pleural Tumours*, Springer-Verlag, Heidelberg, 1999.
- [33] R. Ullmann, A. Schwendel, H. Klemen, G. Wolf, I. Petersen and H.H. Popper, Unbalanced chromosomal aberrations in neuroendocrine lung tumors as detected by comparative genomic hybridization, *Hum. Pathol.* **29** (1998), 1145–1159.
- [34] A. von Deimling, D.N. Louis, K. von Ammon, I. Petersen, T. Hoell, R.Y. Chung, R.L. Martuza, D.A. Schoenfeld, M.G. Yasargil and O.D. Wiestler, Association of epidermal growth factor receptor gene amplification with loss of chromosome 10 in human glioblastoma multiforme, *J. Neurosurg.* **77** (1992), 295–301.
- [35] J. Whang-Peng, C.S. Kao-Shan, E.C. Lee, P.A. Bunn, D.N. Carney, A.F. Gazdar and J.D. Minna, Specific chromosome defect associated with human small-cell lung cancer; deletion 3p(14–23), *Science* **215** (1982), 181–182.
- [36] W. Wick, I. Petersen, R.K. Schmutzler, B. Wolfarth, D. Lenartz, E. Bierhoff, J. Hümmerich, D.J. Müller, A.P. Stangl, J. Schramm, O.D. Wiestler and A. von Deimling, Evidence for a novel tumor suppressor gene on chromosome 15 associated with progression to a metastatic stage in breast cancer, *Oncogene* **12** (1996), 973–978.